

ABSTRACT

Demographic information of an Internet user is predicted based on an analysis of accessed web pages. Web pages accessed by the Internet user are detected and mapped to a user path vector which is converted to a normalized weighted user path vector. A centroid vector identifies web page access patterns of users with a shared user profile attribute. The user profile attribute is assigned to the Internet user based on a comparison of the vectors. Bias values are also assigned to a set of web pages and a user profile attribute can be predicted for an Internet user based on the bias values of web pages accessed by the user. User attributes can also be predicted based on the results of an expectation maximization process. Demographic information can be predicted based on the combined results of a vector comparison, bias determination, or expectation maximization process.